DOCUMENT RESUME

ED 075 506 TM 002 609

AUTHOR Thomas, Hollie B.
TITLE A Case for Using Factor Scores Rather than Summative Scores in Educational Research.
PUB DATE 73
NOTE 13p.; Paper presented at the Joint Session of the Annual Meeting of National Council on Measurement in Education and American Education Research Association (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Comparative Analysis; Educational Research; *Factor Analysis; Grade 9; Rating Scales; *Scores; Suburban Schools; *Test Interpretation; Work Attitudes
IDENTIFIERS *Work Values Inventory; WVI

ABSTRACT

A study designed to investigate the validity of using scale scores in lieu of factor scores for factors derived from a factor analysis technique is presented. The Work Values Inventory (WVI) was selected for use because of its purported factor structure. The population consisted of ninth-grade students from a suburban area of a large metropolitan center. A random sample of approximately 60 students was obtained from each of four schools. Scores for each subject were computed by two different methods: factor scoring and summative scoring. Fewer than 15 factors are required to define the work value dimensions included in the WVI. The common variance between the two techniques ranged from 43.6% to 70.6% with an average of 54.76%. It is tentatively concluded that the degree to which the correlation of summative scores and factors scores computed on the same data deviate from unity is an estimation of their deviation from "conceptual validity." (CK)

27.18

A Case for Using Factor Scores Rather than Summative Scores in Educational Research*

Hollie B. Thomas
University of Illinois

A typical procedure in educational research is to develop a data collecting instrument that is presumably designed to measure an overall variable without considering the underlying constructs. The instrument thus constructed is employed to collect data. Having collected the data the researcher proceeds to "do a factor analysis" without any presuppositions regarding the factorial complexity of the instrument. With the factor analysis in hand, however, it is easy to rationalize the relationship of items and to name the factors derived. The researcher must now make the decision as to what type of scores he will use in the analysis to test the hypotheses he originally constructed only now adding sub-hypotheses for the newly found factors which are tacitly assumed to have factorial validity.

Although these procedures and assumptions appear to the exploratory statistician to be somewhat sound in practice, once the constructs are identified through factor analysis, only those items which load highest on each factor of a rotated factor matrix are included in the computation of a particular "factor score." In addition, in order to simplify the scoring of the instrument, it is a common practice to proceed without theoretical justification to generate "factor scores," or more precisely, scale scores, by summing the standard or raw scores of subjects on those variables which have largest loadings on the rotated factor matrix. As noted by Glass and Taylor (1966) sometimes, though less frequently, the loadings of variables on a factor are used as coefficients in a linear combination of standard scores to yield "factor scores." Glass and Taylor further state that "both of these techniques are

---

* A paper presented to a joint session of the annual meeting of NCME and AERA, New Orleans, 1973.

wrong, whether they are applied to components or in common-factor analysis." Glass and Maguire (1966) note that this procedure will produce a set of "factor scores" having few properties of the legitimate factor scores desired. "Factor scores" thus derived frequently have large non-zero intercorrelations.

Objective of the Inquiry.

This study was designed to investigate the validity of using scale scores in lieu of factor scores for factors derived from a factor analysis technique. More specifically, do scale scores derived by summing the response score of the items which load the highest on the various factors have adequate relationships to the factor scores and produce sufficiently similar results when employed as independent variables to justify their use for attitudinal type research.

Instrumentation.

The Work Values Inventory (WVI) by Super (1970) was selected for use in this study because of its purported factor structure (O'Connor and Kinnane, 1961; Gable and Pruzek, 1971). The WVI is designed as a wide range value inventory to assess the goals which motivate man, including values which are both extrinsic to as well as intrinsic in work. Work values purportedly measured include Creativity, Management, Achievement, Surroundings, Supervisory Relations, Way of Life, Security, Associates, Esthetics, Prestige, Independence, Variety, Economic Return, Altruism, and Intellectual Stimulation. Thus the underlying constructs have been thoroughly considered in the development of this instrument.

Population.

The population consisted of ninth-grade students from a suburban area of a large metropolitan center. A random sample was obtained by randomly ordering the

schools and contacting them in order until four cooperating schools were secured. Five schools were contacted in order to obtain the cooperation of four. A random sample of approximately 60 students was obtained from each school. The instrument was administered to the students during school hours, thus assuring participation of those selected.

Methods.

Scores for each subject were computed by two different methods: factor scores computed from a component analysis and summative scores of the three items loading the highest on each factor. In the latter case, items were assigned to only one scale based on an *a priori* decision to include only items with positive loadings and to assign items not of unit complexity such that when all scales were considered, the highest possible combinations of the loadings were acquired. While many of the items in the first three scales were not of unit complexity, it was necessary to apply this rule for only one item. All other items not of unit complexity loaded among the highest three on only one factor. The factor scores were obtained by first computing a principal components factor analysis followed by a varimax rotation of factors with eigenvalues greater than one. Factor scores were then computed for each subject by multiplying his standard score matrix by the inverse of the correlation matrix by the varimax matrix. This may be represented by the formula:

$$F = ZR^{-1} V \quad \text{(Horn, 1965)}.$$

This method of computing factor scores, commonly referred to as "Little Jiffy" (Kaiser, 1970) was selected for this study because of its popularity in educational research (Cronbach, 1970; Kaiser, 1970, Glass and Taylor, 1966). In addition, there was no reason to assume that any of the variables being analyzed were

operating as random variables. The question of the proper number of factors to extract, a question which has not been answered to date (Tatsuoka, 1971) will not be approached in this paper. Analyses of the scores thus obtained included the computation of the correlation coefficients and factorial analysis of variance utilizing each of the scores as a dependent variable. The independent variables were race, social position, and sex each having two levels. Social position was obtained by employing Hollingshead's (1958) scale. Subjects whose families rated 6 or higher on the scale were in the low social position group.

Results and Discussion.

The results will be presented in three sections: 1) factor analysis, 2) correlations of factor scores and scale scores, 3) an analysis of variance of scores obtained by each method.

Factor Analysis.

The principal components factor analysis of the data from the sample of 234 ninth grade students produced twelve factors with eigenvalues greater than one. Factor loadings resulting from the varimax rotation are presented in Table 1. The highest three loadings on each scale and thus the items used in the calculation of the summative scale scores are underlined. A comparison of the items employed to obtain the scale scores in this study and those recommended by the instrument developed are provided in Table 2. It appears that fewer than the fifteen factors described by Super are required to define the work value dimensions included in the WVI. Gable and Pruzek (1971) found ten factors produced by an image analysis to be meaningful. Items from some of the WVI scales could be appropriately grouped on other scales. Scales which were merged were: 1) Economic Return-Surroundings-Associates, 2) Esthetics-Creativity, and 3) Prestige-Management. Results of the

TABLE 1.
Principal Components Rotated (Varimax) Factor Loadings for Work Value Items*

WVI Items	Factors											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	-09	11	42	17	24	-29	23	-16	18	24	-18	-22
2	-05	-01	80	-01	04	-02	04	-03	05	03	03	02
3	09	61	-16	20	-08	03	-09	30	-00	10	23	-09
4	-03	12	04	14	-02	02	07	09	-05	70	16	01
5	-07	12	05	16	15	66	05	08	07	05	13	09
6	11	49	03	29	03	27	-07	18	13	03	-18	-05
7	02	22	-05	56	03	15	-14	-06	25	06	-09	08
8	19	23	09	-06	-16	19	-39	06	22	34	-14	19
9	23	65	06	06	16	06	01	10	-02	-01	06	12
10	-02	02	39	12	27	19	09	-02	49	10	12	-03
11	01	26	03	05	10	23	-12	02	11	13	66	-18
12	13	40	13	13	45	36	-03	-10	08	-00	13	-13
13	23	07	04	13	69	08	07	05	-08	-04	10	-06
14	18	22	-08	06	01	01	-12	70	10	13	04	-20
15	03	03	22	59	18	-10	26	07	11	23	07	-05
16	02	01	05	77	19	15	07	15	-12	06	-02	15
17	07	08	19	18	70	02	08	-05	17	09	15	25
18	16	21	09	-13	32	03	-01	06	-07	02	61	-12
19	22	74	07	-04	11	-02	-04	-04	05	-08	13	02
20	02	07	20	33	09	19	-11	-07	15	02	-09	65
21	-14	25	17	13	30	35	19	10	34	04	-13	21
22	26	59	-12	-03	06	24	20	17	-04	21	10	15
23	06	-04	07	00	13	13	58	20	04	35	-16	12
24	06	11	12	05	-00	07	32	71	13	02	00	17
25	32	42	02	-12	18	-16	12	16	-11	17	08	41
26	25	01	04	-09	15	42	07	-13	19	34	14	24
27	44	16	34	-05	05	08	21	10	-04	18	11	02
28	65	27	10	05	29	-08	-08	10	06	00	-09	-11
29	45	-14	00	10	12	08	06	-00	15	57	-11	-08
30	42	-13	63	09	13	12	08	08	-12	04	00	18
31	41	-02	62	20	09	19	12	10	-10	-05	05	17
32	46	-04	10	19	-08	05	04	05	00	41	06	32
33	63	22	-01	-06	05	01	-20	32	09	01	-18	01
34	65	31	19	10	-07	14	00	06	-13	09	03	-05
35	36	19	10	12	-15	53	29	-02	09	02	06	-09
36	57	27	-05	-02	12	-06	21	09	02	06	18	15
37	45	16	12	09	-01	-34	02	39	24	-07	07	10
38	24	13	27	14	02	14	60	02	13	00	-02	-09
39	47	51	-14	-02	-14	13	34	-02	09	15	17	00
40	13	-00	-12	07	-04	09	04	31	74	00	00	06
41	30	25	-05	40	-23	-15	05	-02	39	01	14	36
42	42	48	-10	10	-23	-07	08	-00	13	-09	35	04
43	62	09	01	-03	06	07	14	-04	03	08	44	11
44	56	-07	06	12	30	10	15	-07	36	03	05	07
45	40	-09	24	45	10	06	37	-04	12	02	00	-00

* decimals have been omitted.

TABLE 2.
List of Items Employed to Obtain Cumulative Scores

WVI Scale	WVI Items	Factor Number	Items Correlating Highest with Scale (groups by Varimax Rotation)
1. Creativity	15, 16, 45	IV	7, 15, 16
2. Management	14, 21, 37	VIII	14, 24, 37
3. Achievement	13, 17, 44	V	12, 13, 17
4. Surroundings	12, 25, 36		
5. Supervisory Relations	11, 18, 43	XI	11, 18, 43
6. Way of Life	10, 26, 35	VI	5, 26, 35
7. Security	9, 19, 42	II	3, 9, 19
8. Associates	8, 27, 34		
9. Esthetics	7, 20, 41	XII	20, 25, 41
10. Prestige	6, 28, 33	I	28, 33, 34
11. Independence	5, 21, 40	IX	10, 40, 44
12. Variety	4, 29, 32	X	4, 29, 32
13. Economic Return	3, 22, 39		
14. Altruism	2, 30, 31	III	2, 30, 31
15. Intellectual Stimulation	1, 23, 38	VII	23, 38, 45

present study were similar in that items on the scales of Associates [illegible] with Prestige. Items from the scale of Economic Return, Security and Satisfaction were also grouped.

Correlations.

The problems that arise from computing scale scores from the highest loading items of a factor analysis can be observed in their intercorrelations. These scale scores, often assumed to be appropriate estimates of the factor scores, are frequently highly intercorrelated. The intercorrelations, for the summative scores shown in Table 3 computed using the three items that loaded the highest on the rotated factors, ranged up to .47.

A second problem with summative scale scores is their modest correlation with the actual varimax factor score from which the scales were derived. As shown in Table 4 on the diagonal of the matrix, the correlation of the actual varimax factor score and the corresponding scale score ranged from .66 to .84. Thus, the common variance between the two techniques ranged from 43.6% to 70.6% with an average of 54.76%.

A Research Example.

Researchers may be able to accept scale scores which correlate with the factor scores as high as those found in this study on the basis of the results of the correlational analysis. The true test, however, is to employ both types of scores in a typical educational research analysis and to ascertain what types of conclusions would be made from each. Thus an analysis of the two sets of scores was made, utilizing the same data that was factor analyzed to explore the effects of sex, race, and social position on work values. This 2x2x2 factorial analysis of variance

TABLE 3
Intercorrelations of Summative Scale Scores

Scale	Scale 1	2	3	4	5	6	7	8	9	10	11	12
1.	1.00											
2.	0.42	1.00										
3.	0.31	0.04	1.00									
4.	0.16	0.17	0.22	1.00								
5.	0.30	0.31	0.34	0.34	1.00							
6.	0.27	0.23	0.26	0.26	0.36	1.00						
7.	0.27	0.06	0.43	0.33	0.31	0.34	1.00					
8.	0.47	0.34	0.17	0.21	0.12	0.13	0.26	1.00				
9.	0.32	0.17	0.28	0.31	0.36	0.30	0.28	0.34	1.00			
10.	0.34	0.13	0.30	0.27	0.18	0.35	0.37	0.21	0.35	1.00		
11.	0.37	0.40	0.21	0.09	0.42	0.30	0.19	0.23	0.25	0.24	1.00	
12.	0.37	0.34	0.26	0.37	0.25	0.29	0.28	0.32	0.30	0.30	0.29	1.00

TABLE 4.

Cross-Correlations of Actual Varimax Factors with Summative Scale Scores

Actual Varimax Factor	Summative Scale Scores											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
I.	0.81	0.15	0.34	0.04	0.19	0.27	0.31	0.32	0.34	0.40	0.39	0.31
II.	0.34	0.84	-0.07	0.14	0.24	0.15	-0.00	0.22	-0.00	-0.03	0.24	0.33
III.	0.11	-0.01	0.83	0.10	0.17	0.09	0.25	0.08	0.11	0.08	0.05	0.00
IV.	0.03	0.09	0.12	0.80	0.18	0.07	0.25	0.06	0.12	0.17	-0.04	0.26
V.	0.12	0.07	0.11	0.15	0.80	0.08	0.12	-0.00	0.21	0.01	0.22	0.39
VI.	0.04	0.02	0.13	0.10	0.19	0.74	0.13	-0.13	0.16	0.03	0.16	-0.07
VII.	-0.13	-0.06	0.09	0.05	0.04	0.17	0.69	0.02	0.13	0.02	0.01	0.03
VIII.	0.21	0.15	0.03	0.03	-0.05	-0.06	0.09	0.81	0.12	0.05	0.03	0.02
IX.	0.01	0.02	-0.07	0.12	0.07	0.17	0.12	0.21	0.76	0.05	0.05	0.21
X.	0.03	0.01	0.01	0.14	0.02	0.21	0.13	0.04	0.05	0.76	0.12	0.03
XI.	-0.10	0.20	0.04	-0.01	0.17	0.13	-0.09	0.07	0.08	0.07	0.75	0.09
XII.	-0.07	-0.01	0.17	0.03	0.02	0.13	0.02	0.02	0.06	0.10	0.04	0.69

produced considerably divergent results for the two sets of scores. As shown in Table 5, the use of the summative scores produced a greater number of significant main effects than were obtained when the factor scores were employed. However, some significant main effects were common to both procedures, but each procedure produced unique significant main effects.

The researcher interpreting the analysis of the summative scale scores would conclude that females are more altruistic and esthetic than males; whites place more value on work that will provide variety, intellectual stimulation, and opportunities to provide service to their fellow man (altruism) than blacks; and that students from higher social position families are more altruistic than are students from low social position families.

If the same researcher had chosen to employ factor scores, he would have concluded that females are more altruistic than males; whites value variety in their work and seek prestige more than do blacks; and that students from high social position families place more value on independence - way of life and less value on supervisory relations than do students from low social position families. Thus the two methods of scoring produced consistent significant main effects in these analyses for race on the variety scale and sex on the altruism scale. All other significant main effects were unique to the particular scoring procedure.

Conclusions.

Some proponents of factor analysis, such as Guertin and Bailey (1970) contend that the paradigm for determining validity by correlating factors on an instrument with some external criterion is antiquated. The paradigm proposed is to employ factor analysis to "conceptually validate" new instruments. Many researchers have employed this paradigm to "validate" research instruments, but have not followed

TABLE 5.
Summative Scale Score and Factor Score Means by Sex, Race, and Social Position

Work Value		Sex			Race			Social Position		
		Male	Female	Significance of Main Effect	Black	White	Significance of Main Effect	Low	High	Significance of Main Effect
Associates - Prestige	SS	10.6	10.7	N.S.	10.4	11.0	N.S.	10.4	11.0	N.S.
	FS	- .19	- .08	N.S.	- .38	.11	.01	- .25	- .02	N.S.
Security - Economic Ret.- Surroundings	SS	12.4	12.4	N.S.	12.5	12.3	N.S.	12.3	12.4	N.S.
	FS	.20	.10	N.S	.25	.05	N.S.	.11	.19	N.S.
Altruism	SS	11.0	12.3	.001	11.1	12.2	.01	11.2	12.1	.05
	FS	- .39	.09	.01	- .30	.00	N.S.	- .21	- .01	N.S.
Creativity - Esthetics	SS	10.0	11.0	N.S.	10.5	10.5	N.S.	10.8	10.2	N.S.
	FS	- .18	.13	N.S.	- .03	.03	N.S.	.11	- .16	N.S.
Achievement -	SS	11.8	12.0	N.S.	11.8	12.1	N.S.	11.9	11.9	N.S.
	FS	- .15	- .10	N.S.	- .12	- .13	N.S.	- .06	- .19	N.S.
Independence- Way of Life	SS	12.3	12.8	N.S.	12.2	12.8	N.S.	12.2	12.8	N.S.
	FS	- .11	- .12	N.S.	- .22	- .02	N.S.	- .46	.22	.001
Intellectual Stimulation	SS	10.9	11.5	N.S.	10.5	11.9	.001	10.9	11.6	N.S.
	FS	- .14	.01	N.S.	- .15	.02	N.S.	- .17	.03	N.S.
Management	SS	9.8	9.3	N.S.	9.7	9.5	N.S.	9.7	9.4	N.S.
	FS	.16	- .17	N.S.	.08	- .10	N.S.	- .01	.00	N.S.
Independence	SS	11.4	11.8	N.S.	11.6	11.6	N.S.	11.5	11.8	N.S.
	FS	- .05	.16	N.S.	.21	.00	N.S.	.16	.06	N.S.
Variety	SS	9.8	10.3	N.S.	9.4	10.7	.01	10.0	10.2	N.S.
	FS	- .11	- .13	N.S.	- .32	.08	.05	- .06	- .18	N.S.
Supervisory Relations	SS	11.6	12.0	N.S.	11.7	11.9	N.S.	12.0	11.7	N.S.
	FS	.12	.30	N.S.	.34	.07	N.S.	.40	.02	.05
Esthetics	SS	9.7	10.7	.05	9.8	10.6	N.S.	10.6	9.8	N.S.
	FS	- .16	.16	N.S.	- .08	.08	N.S.	.15	- .15	N.S.

next logical step to produce factor scores which would yield individual subjects on the underlying constructs which have been "validated." Rather they procede to sum the raw scores of the highest loading items which is often throwing together junk with the score being sought. In the research example, it was shown that the two methods of scoring produce substantially different results. An investigation of the intercorrelations of summative scale scores as well as the cross-correlations of the summative scale scores with the actual varimax factor scores reveal that the Associate-Prestige value, the most salient factor, has a considerable amount of common variance with at least six other scales. This common variance could easily account for significant main effects in the analysis of variance that emerged for the summative scale scores and not for the factor scores.

Since varimax factor scores, as all other factor scores in component analysis, are uncorrelated, it seems reasonable to assume that the degree to which the correlation of summative scores and factor scores computed on the same data deviate from unity, is an estimation of their deviation from "conceptual validity."

Bibliography.

Cronbach, L. J. Essentials of psychological testing (3rd Edition). New York: Harper and Row, 1970.

Gable, Robert K., & Pruzek, Robert M. Super's work values inventory: two multivariate studies of interitem relationships. The Journal of Experimental Education, 1971, 40, 41-50.

Glass, Gene V., & Maguire, Thomas O. Abuses of factor scores. American Education Research Journal, 1966, 3, 297-304.

Glass, Gene V., & Taylor, Peter A. Factor analytic methodology. Review of Educational Research, 1966, 36, 565 - 587.

Guertin, Wilson H. & Bailey, John P. Introduction to modern factor analysis. Ann Arbor: Edwards Brothers Inc., 1970.

Horn, John L. An empirical comparison of methods for estimating factor scores. Educational and Psychological Measurement, 1965, 25, 313-322.

Kaiser, Henry F. A second generation little jiffy. Psychometrika, 1970, 35, 401-415.

O'Connor, James P., & Kinnane, John F. A factor analysis of work values. Journal of Counseling Psychology, 1961, 8, 263-267.

Super, Donald E. Work values inventory. Boston: Houghton Mifflin Co., 1970.

Tatsuoka, Maurici M. Multivariate analysis techniques for educational and psychological research. New York: Wiley & Sons, 1971.